Midterm - CSC 262 - October 15, 2015

NAME: ____________________________________________________________

This exam is closed book. You are allowed one aid sheet on a standard $8.5 \times 11$ inch paper (both sides) and a calculator. Answer the questions in the space provided. You have 1-1/4 hours. Answer all four questions. All questions have equal weight. You are encouraged to read each question completely before starting.

1. An urn contains 20 balls. Exactly two are labeled by number $i$, for $i = 1, \ldots, 10$. A random sample (without replacement) of 4 balls is taken. What is the probability that no number is represented twice?

SOLUTION It will be useful to temporarily label each numbered pair red and green, so that we have 20 distinct balls. We have

$$D = \binom{20}{4} = \frac{20 \times 19 \times 18 \times 17}{4!} = 4845$$

possible selections. Use the rule of product:

1. Select 4 unique labels from 10, $n_1 = \binom{10}{4} = \frac{10 \times 9 \times 8 \times 7}{4!} = 210$.
2. Select a color for each label, $n_2 = 2^4 = 16$.

There are

$$N = n_1 \times n_2 = 210 \times 16 = 3360$$

such selections, so

$$P(\text{ No number represetned twice }) = \frac{3360}{4845} \approx 0.693.$$

2. A random variable $X$ possesses the following density function for some constant $c$:

$$f_X(x) = \begin{cases} cx^3 & ; \quad x \in [0,3] \\ 0 & ; \quad otherwise \end{cases}.$$

(a) Determine $c$.

(b) Determine the 0.5-quantile for this density.

(a) The integral of a density evaluates to 1, so

$$1 = \int_0^3 cx^3 dx = cx^4/4\big|_0^3 = c \times 81/4,$$

giving $c = 4/81$.

(b) The CDF is

$$F_X(x) = \begin{cases} 0 & ; \quad x < 0 \\ x^4/81 & ; \quad x \in [0,3) \\ 1 & ; \quad x \geq 3 \end{cases}.$$

The 0.5-quantile $q$ is the solution to

$$0.5 = F_X(q) = q^4/81, \quad \text{or} \quad q = (0.5 \times 81)^{1/4} \approx 2.523.$$

3. Let $X_1$ and $X_2$ be independent geometric random variables with means $1/p_1$ and $1/p_2$ respectively.

   (a) If $Y = \min(X_1, X_2)$, show that $Y$ is a geometric random variable with mean $1/\alpha$, where $\alpha = 1 - (1-p_1)(1-p_2)$.

   (b) Prove the following equality:

$$P(X_1 > X_2) = \frac{(1-p_1)p_2}{1-(1-p_1)(1-p_2)}.$$

   (a) The CDF of $X \sim geom(p)$ is $F_X(k) = 1-(1-p)^k$, and the PMF is $p_X(k) = p(1-p)^{k-1}$, $k \geq 1$. We have

$$\begin{aligned} P(Y > k) &= P(X_1 > k \text{ and } X_2 > k) \\ &= P(X_1 > k)(X_2 > k) \\ &= (1-p_1)^k(1-p_2)^k \\ &= [(1-p_1)(1-p_2)]^k, \end{aligned}$$

   that is, $Y \sim geom(1-(1-p_1)(1-p_2))$.

   (b) Recall geometric series, for $r < 1$,

$$\sum_{i=0}^{\infty} r^i = \frac{1}{1-r}$$

   We have, by independence

$$P(X_1 > X_2 \mid X_2 = k) = P(X_1 > k).$$

   By the law of total probability, conditioning on partition $A_k = \{X_2 = k\}$, $k = 1, 2, \ldots$, we have

$$\begin{aligned} P(X_1 > X_2) &= \sum_{k=1}^{\infty} P(X_1 > X_2 \mid X_2 = k)P(X_2 = k) \\ &= \sum_{k=1}^{\infty} P(X_1 > k)P(X_2 = k) \\ &= \sum_{k=1}^{\infty}(1-p_1)^k p_2(1-p_2)^{k-1} \\ &= (1-p_1)p_2\sum_{k=0}^{\infty}[(1-p_1)(1-p_2)]^k \\ &= \frac{(1-p_1)p_2}{1-(1-p_1)(1-p_2)}. \end{aligned}$$

4. A test for the presence of an infection is developed. It is administered to subjects whose infection status is known (83 are infected, 420 are not infected). The results are summarized in the following contingency table. Calculate the *positive predictive value* (PPV) and the *negative predictive value* (NPV) of the test for a population with an infection prevalence of $prev = 0.05$.

| | | Infection | | |
|---|---|---|---|---|
| | | Positive | Negative | Total |
| Diagnostic Test | Positive | 72 | 5 | 77 |
| | Negative | 11 | 415 | 426 |
| | Total | 83 | 420 | 503 |

SOLUTION We have

$$\begin{aligned} sens &= \frac{TP}{TP+FN} = \frac{72}{83} \approx 0.867 \\ spec &= \frac{TN}{TN+FP} = \frac{415}{420} \approx 0.988. \end{aligned}$$

For specific prevalence $prev = 0.05$ we have

$$\begin{aligned} PPV &= \frac{sens \times prev}{sens \times prev + (1 - spec) \times (1 - prev)} \\ &= \frac{0.867 \times 0.05}{0.867 \times 0.05 + (1 - 0.988) \times (1 - 0.05)} \\ &\approx 0.793 \end{aligned}$$

and

$$\begin{aligned} NPV &= \frac{spec \times (1 - prev)}{spec \times (1 - prev) + (1 - sens) \times prev} \\ &= \frac{0.988 \times (1 - 0.05)}{0.988 \times (1 - 0.05) + (1 - 0.867) \times 0.05} \\ &\approx 0.993. \end{aligned}$$